CERES Data Management Group Status Report

30th CERES Science Team Meeting

September 10th, 2018

Jonathan Gleason
Kathleen Moore Dejwakh

Data Management Team

Systems/ Optimization:
Nelson Hillyer
Mike Linsinbigler

Product Web Support:
Churngwei Chu
Liz Heckert
Ed Kizer
Cristian Mitrescu

Config. Mgnt & Production Support:
Tammy Ayers
Joanne Saunders
Dennis Keyes
Carla Grune
Jeremie Lande

DMT Leadership:
Jonathan Gleason
Kathleen Moore Dejwakh
Walt Miller

Instrument:
Denise Cooper
Thomas Grepiotis
Dianne Snyder
Mark Timcoe

ERBE-like:
Dale Walikainen

Clouds:
Sunny Sun-Mack
Yan Chen
Rita Smith
Ricky Brown

Convolution:
Igor Antropov

Inversion:
Victor Sothcott

SARB:
Tom Caldwell

TISA:
Josh Wilkins
Moguo Sun
Cathy Nguyen

Overview

- Recent DMT Activity
- Data Product Availability
- Reprocessing Activities
- Code Improvement Initiatives
- Systems Status
- Coming improvements

DMT Recent Activity

Recent Activity

45 Software & data deliveries since May 15th, 2018

- **CATALYST (7)** – Mostly PGE module updates
- **PR Tool (4)** – Database updates
- **PerlLib (2)** – Library maintenance
- **Archiver Package (2)** – New archiver scripts
- **Coefficients and Ancillary File Deliveries (20)**
  - Gains, SRFs, MATCH, TSI & GGEO
- **Instrument (1)** – Script Updates
- **ERBElike (1)** – Last PGE migrated off P6!!!
- **Clouds (1)** – MODIS C6.1 update
- **Inversion (1)** – Script Update
- **SARB (1)** – Bug fix
- **Tisa (5)** – Bug fixes

Data Availability

Edition 4 Product Availability

Product	Platform	Processed Through	Current	Publicly Available
BDS	Terra & Aqua	May 31, 2018	Yes	Yes
SSF*	Terra & Aqua	May 31, 2018*	Yes	Yes
SSF1deg-Hour*	Terra & Aqua	April 30, 2018*	Yes	Yes
SSF1deg-Day/Month*	Terra & Aqua	April 2018*	Yes	Yes
SYN1deg**	Terra + Aqua	March 2018	Yes	Yes
CldTypHist	Terra + Aqua + Geo	February 28, 2017	February 2019	No

*SSF & SSF1deg reprocessed and released in Summer 2018 using MODIS C6.1 between March 2016 and March 2018

**SYN1deg to be reprocessed and re-released Fall 2018

S-NPP FM5 Product Availability
Edition 1

Product	Platform	Processed Through	Current?	Publicly Available
BDS	S-NPP	April 30, 2018	Yes	Yes
SSF		April 30, 2018	Yes	Yes
SSF1deg-Hour		March 31, 2018	Yes	Yes
SSF1deg-Day/Month		March 31, 2018	Yes	Yes
SYN1deg	S-NPP + Terra	November 2017	March 2019	Yes

NOAA-20 FM6 Product Availability
Edition 1-CV

Product	Platform	Processed Through	Current?	Publicly Available
BDS	NOAA-20	August 31, 2018	Yes	Yes
ES8		August 31, 2018	Yes	Yes
ES4 / ES9		July 31, 2018	Yes	Yes

Reprocessing Activities

Edition 4 Reprocessing Campaign

New MODIS Collection 6.1

- Reprocess SSF & SSF1deg products Mar 2016 - Mar 2018
- Reproduce internal MATCH aerosol product for entire record

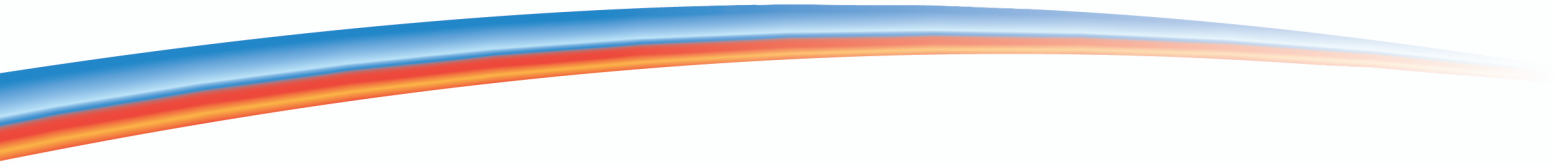
Incidentally discovered McIDAS distribution issues with Himawari 8 offline processed Geo clouds

- Reprocess Geo clouds Jul 2015 - March 2018

Reprocess Edition4 SYNdeg products

- Mar 2000 - Jun 2015: Only change is C6.1 MATCH aerosol input (impacts computed fluxes only)
- Jul 2015 - Feb 2016: Corrected Geo clouds and C6.1 MATCH
- Mar 2016 - Mar 2018: New C6.1 derived SSF as input

Reprocess CldTypHist from Jul 2015 - Mar 2018 with C6.1 SSF and corrected Geo



Code Improvement Initiatives

External Code Assessment

Partnering with Booze Allen Hamilton (BAH) to perform independent analysis of CERES production code

- Initial 90-day consulting period with extension option
- BAH brings custom tool that uses machine learning to ID candidate code for targeted improvements

We provide…

- All active CERES PGEs excluding ERBElike
- Metrics from in-house assessment and GProf code profiling results

They deliver…

- Preliminary assessment results of our software
- List of code blocks for improvements and their location in the code
- Statistics that characterize different groups of code and recommendations for how to improve those blocks

Kathleen Moore Dejwakh leading

Multiple Initiatives to Streamline Code

Code optimization / parallelization support from ODU

- Leveraged support from Langley Digital Transformation Initiative
- ODU Computer Science Dept. faculty (Mohammad Zubair) available to advise CERES developers
- August kickoff discussion with Clouds WG (Leo & Geo code) - takeaway: focus on “low hanging fruit”

Continued support to Tisa WG from DMT Leadership

- Initial assessment of averaging code identified code patterns/ practices candidate for improvement
- Results and recommendations shared with team

Kathleen Moore Dejwakh leading both of these initiatives

Systems Status

Systems Supporting CERES

Goal: Homogeneous processing environment

Production:

- Nearly all CERES PGEs execute on x86 processors
- S-NPP Clouds PGEs still run on P7 platform
  - Recently lost ½ of P7 capacity due to HW failure
  - Do not plan to replace

Offline Science Computing Facility (SCF):

- Primarily x86 and P6 blades – (P6 no longer under maintenance, Please migrate!)
- Limited P7 blades available for testing

Production and SCF eventually to migrate to OpenShift on-premise cloud environment

- Websites will migrate to first

New SCF Support Model

ASDC contract support split by ESDIS-funded and SCF-funded tasks starting October 2018

- Clear cost tracking between ESDIS and non-ESDIS funds
- Opportunity to restructure SCF support model at ASDC
  - Will require WYE reporting per project (i.e. CERES, CALIPSO, HSRL, etc)
  - Guiding principle: “First do no harm!”

More clearly defined issue resolution approach

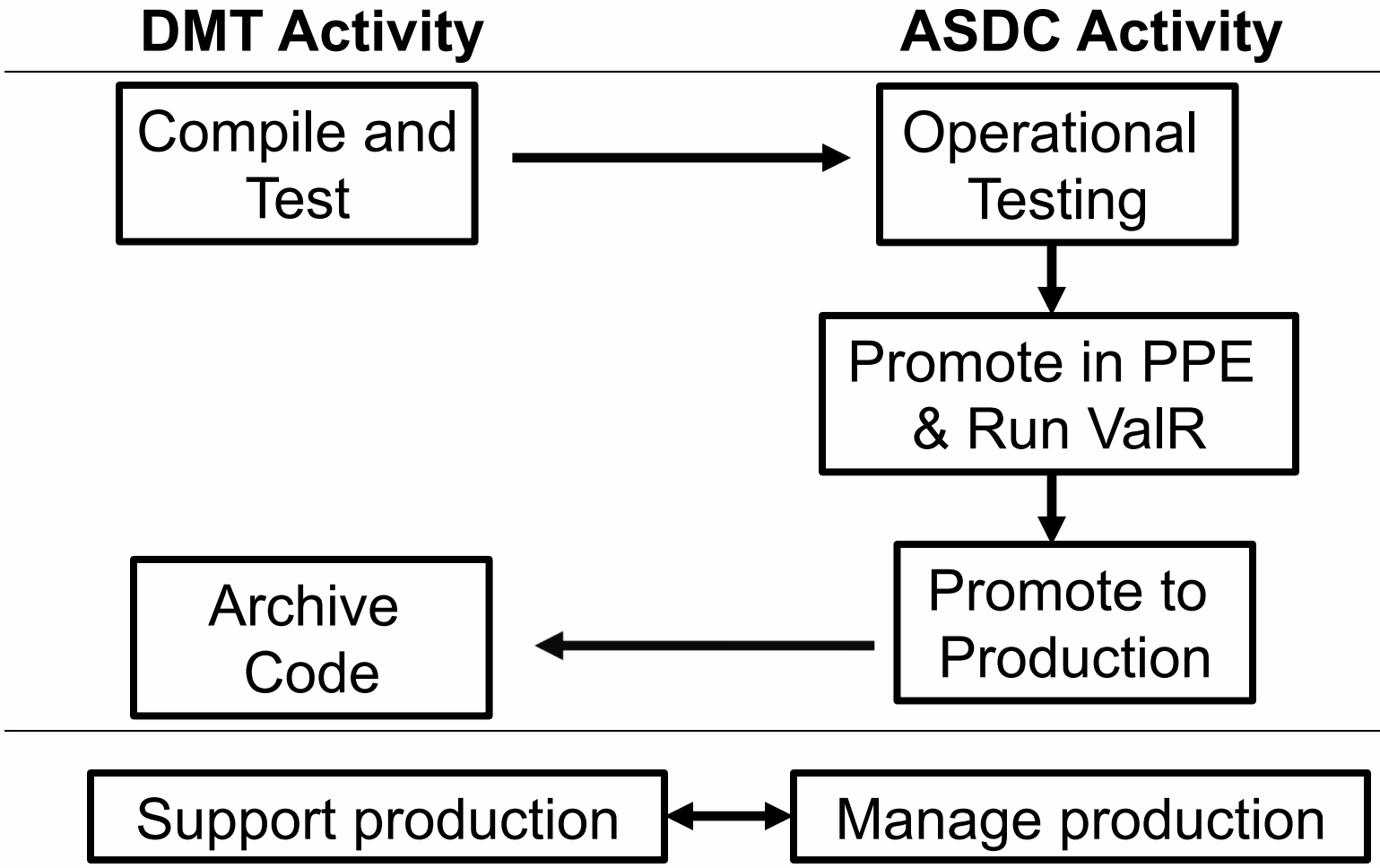
- All support requests to start to IT Service Desk ticket
- Defining Service Level Agreements for SCF support

Process to be a dialog: What do you want to see from ASDC?

- SCF task TPOC: Chris Harris; Alt: Jonathan Gleason

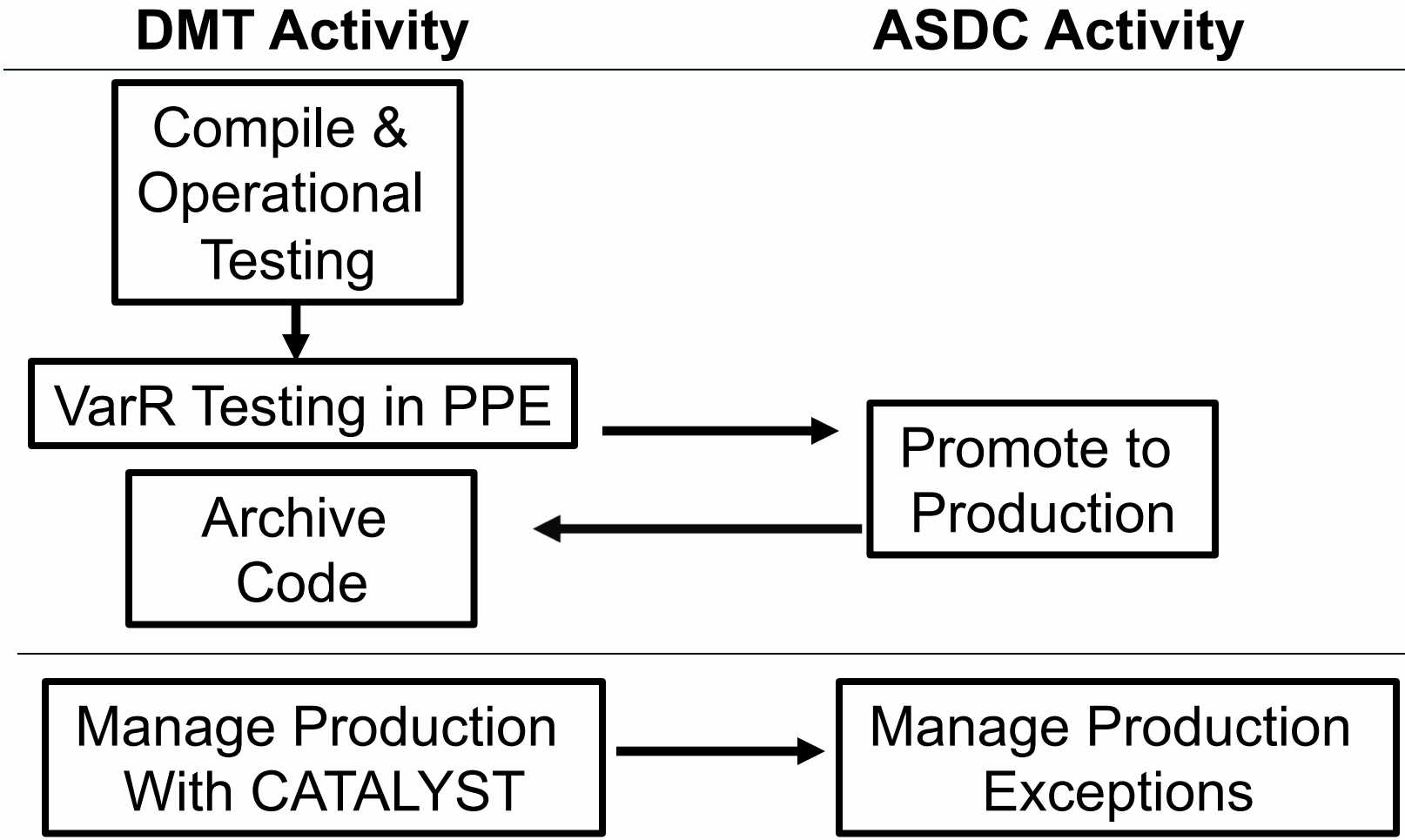
Paradigm Shift – DMT In-House

CATALYST enables significant changes in how we manage production data processing



Paradigm Shift – DMT In-House

CATALYST enables significant changes in how we manage production data processing



New Archive Scripts

Post production file disposition historically done by collection of "epilogue" scripts that archive files and clean disk workspace

Legacy scripts developed pre-mission & maintained by ASDC

- File management database dependency
- Time consuming to update
- Pre-delivery testing difficult and inefficient

DMT implemented new library-based archive scripts

- File management policy implemented via flat config. file (no database dependency & easy to update)
- Streamlined testing includes archive scripts
- Logic leveraged from CATALYST

Resulting Improvements

Example improvements from automated process management with CATALYST utility

- Originally took ~2 years to run Edition 4 (about 13 years) – Now can process up to 3 months per day
- Previously could run 2-3 months of SYNI per week – Now routinely process ~1 year per week
- Forward processing managed by software in the background

DMT “owns” more of the process

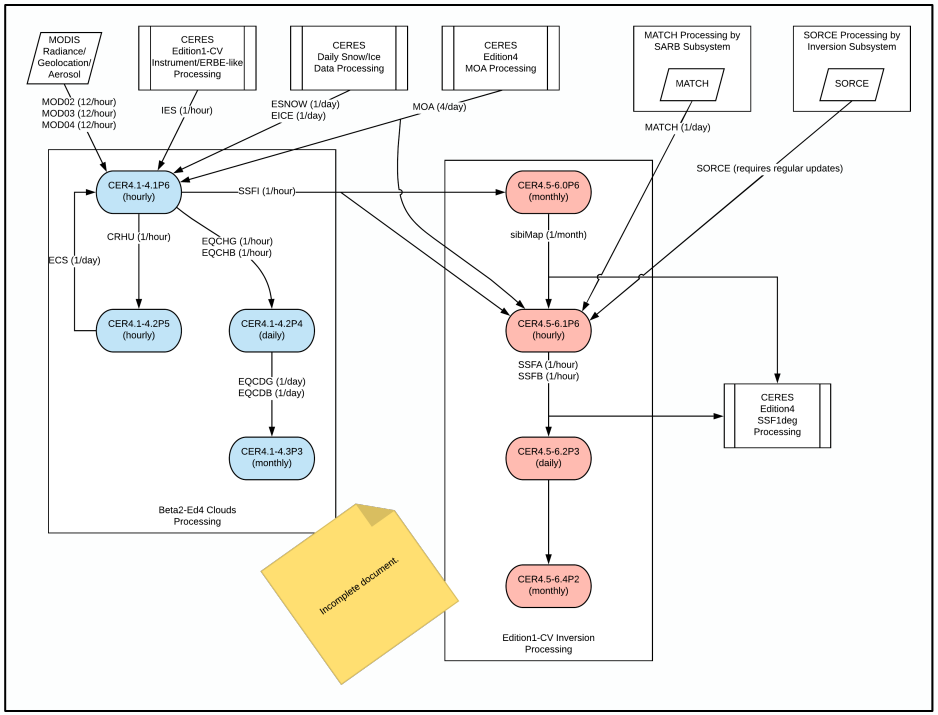
- Delivery and test process significantly streamlined
- Issues resolved quickly by a collocated under one organizational structure (DMT)
- Issue avoidance from improved procedures

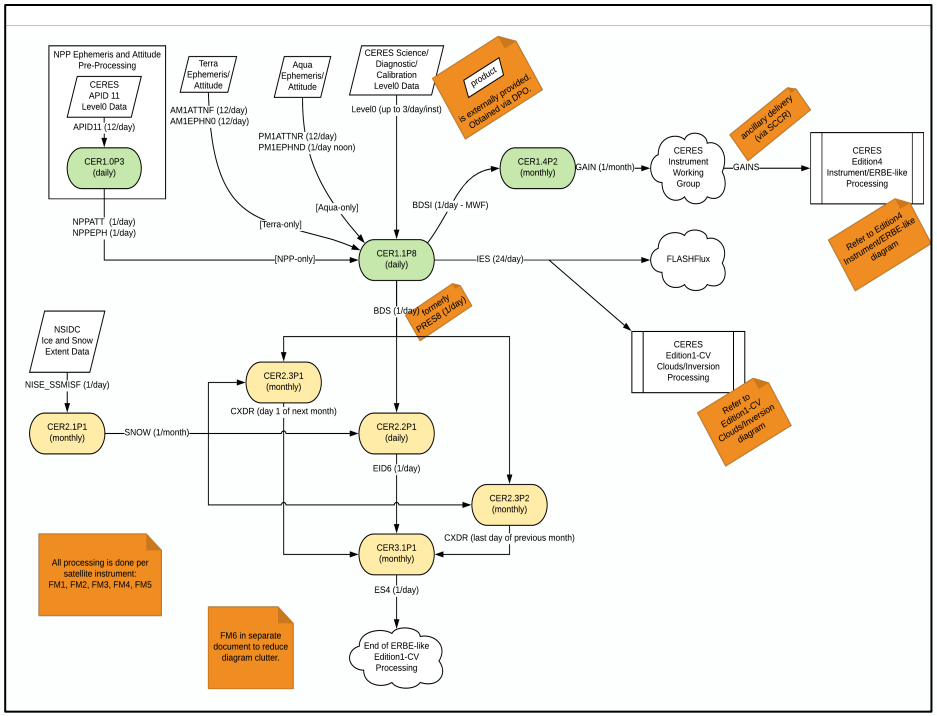
Coming Improvements

Production Documentation

Graphicly document software processing streams

- Processing significantly changed since Edition 2 (last diagrams)
- Dependencies already graphed for CATALYST





Metadata Standardization

Actively evaluating Universal Metadata Model (UMM) standards for use in CERES data products

- UMM-Collection – Metadata common to the collection
- UMM-Granule – specific to a given instance/granule of data
- UMM-Var – Variable level specifics: includes options for recording quality
- UMM-Common – Standard across all

Standardized metadata enables searchability and discoverability from 3rd party tools

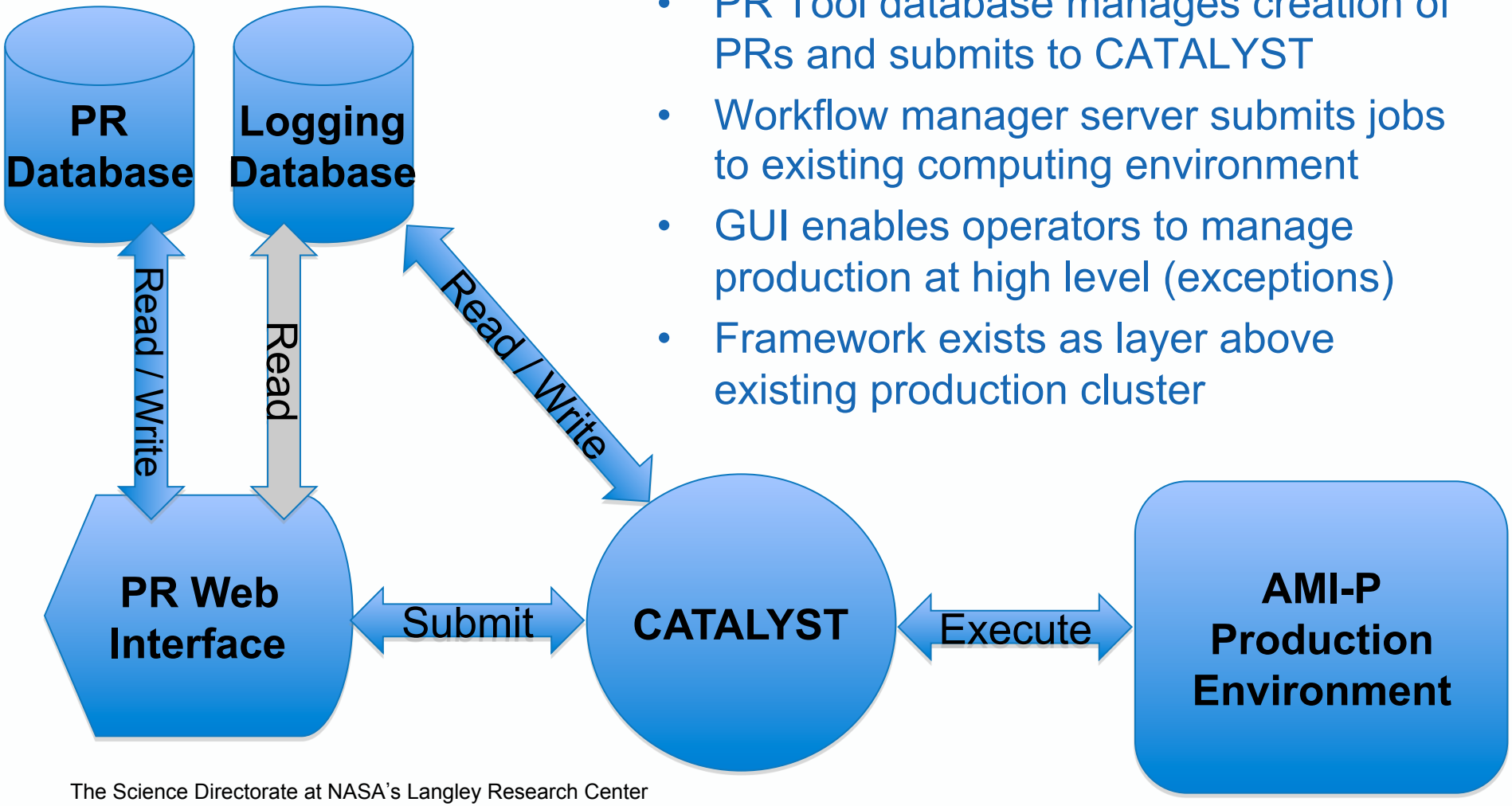
- Have seen several examples with ESDIS Earthdata Search tool where CERES data was not discoverable
- Enables connectivity with tools like OpenDAP, etc

Backup Slides

Automation Framework

CERES AuTomAted job Loading sYSTem (CATALYST)

- PR Tool database manages creation of PRs and submits to CATALYST
- Workflow manager server submits jobs to existing computing environment
- GUI enables operators to manage production at high level (exceptions)
- Framework exists as layer above existing production cluster



CATALYST Status

CATALYST Release 1.0 supported Clouds and Inversion PGEs only

- Processed Beta2-Ed4 SSFs through December 2014 for Terra & Aqua
- Successfully ran ~505,448 jobs (328 data months)

CATALYST Release 2.0

- Implements user requested features for increased usability
- Included significant server improvements for maintainability and sustainability
- Version 2.0 promoted live January 5th, 2016
- Migration of remaining Edition 4 PGEs into CATALYST in progress

Delivery Tracking Using JIRA

CERES Deliveries In JIRA

Now using JIRA to track CERES deliveries and communicate between ASDC and DMT

- SCCR high level workflow – 1 ticket per SCCR
- Each SCCR ticket composed of 1 or more “sub-tickets” of one of 5 different types
  - PGE Delivery - 19 to 25 possible workflow steps
  - Delta Script Delivery
  - Delta Data Delivery
  - Library Delivery (CERESLib / Perl_LIB)
  - Automation Delivery (CATALYST or PR Tool)
- JIRA provides dashboard tool for customized reporting

CERES Deliveries In JIRA

Have successfully used this system for >100 deliveries

- JIRA tickets replace email for communication
- Workflow handles transition of ownership between DMT and ASDC (Ex: when ValR approval needed)
- System open for all with JIRA account to see status of any ticket
- Option exists to create “public” dashboard pages for WG leads to view status for deliveries of interest

Plan to transition ticket ownership to DMT Subsystem Leads in May

- SS Developer will track tickets
- ValR evaluation directly linked to subsystem
- DMT Lead not required middle person

https://asdcjira.larc.nasa.gov:8443/secure/Dashboard.jspa

Production Status Reporting

Production status available in Confluence

- Monthly priorities from DMT
- Weekly status updates from Ops

Click below link to view (Open within LaRC firewall)

- “DMT Production Information”: Links to page of all monthly processing priorities provided
- “OPS Processing Status”: Links to list of months, each month links to page of weekly updates for that month

https://asdcjira.larc.nasa.gov:9443/pages/viewpage.action?pageId=18352593

Started a Spring Science Team Meeting processing priorities dedicated page in January 2016 (time to prepare for Fall Meeting)